

Information infrastructure for Australia's Integrated Marine Observing System

Marton G. Hidas¹ · Roger Proctor¹ · Natalia Atkins¹ · Julian Atkinson¹ · Laurent Besnard¹ · Peter Blain¹ · Philip Bohm¹ · Jon Burgess¹ · Kim Finney¹ · Dan Fruehauf¹ · Guillaume Galibert¹ · Xavier Hoenner¹ · Jacqui Hope¹ · Craig Jones¹ · Sebastien Mancini¹ · Benedicte Pasquer¹ · David Nahodil¹ · Kate Reid¹ · Katherine Tattersall¹

Received: 12 November 2015 / Accepted: 17 May 2016 / Published online: 25 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Australia's Integrated Marine Observing System (IMOS, imos.org.au) is research infrastructure to establish an enduring observing program for Australian oceanic waters and shelf seas. The observations cover physical, biological, and chemical variables to address themes of multi-decadal ocean change, climate variability and weather extremes, boundary currents and inter-basin flows, continental shelf processes and ecosystem responses.

IMOS observations are collected by national facilities based on various platform types and operated by partner institutions around the country. In this paper we describe the infrastructure and workflows developed to manage and distribute the data to the public. We highlight the existing standards and open-source software we have adopted, and the contributions we have made. To demonstrate the value of this infrastructure we provide some illustrations of use and uptake.

All IMOS data are freely and openly available to the public via the Ocean Portal (<https://imos.aodn.org.au>). All IMOS-developed software is open-source and accessible at <https://github.com/aodn>.

Keywords Marine observations · Research infrastructure · Data management · Data access · Public data · Open-source software

Communicated by: H. A. Babaie

✉ Marton G. Hidas
marty.hidas@utas.edu.au

¹ Integrated Marine Observing System, University of Tasmania, Private Bag 110, Hobart, TAS 7001, Australia

Introduction

Ocean observations are important for many modern day requirements, for example as inputs/validation for ocean forecasting/hindcasting, studying climate variability, contributing to environmental assessments and monitoring, commercial fishing, and to aid design of offshore installations.

However, observing the ocean is expensive due to the spatial and temporal extents to be sampled, and the range of variables to measure. While technological advances have made it possible to collect vast quantities of data, managing and distributing these data effectively is a challenge in itself.

Among other benefits, a large-scale collaborative observing program is more efficient than individual research groups as it shares the cost of observations and data management while addressing multiple research goals. The value of observations can be further increased by making the data available to the global research community and the public. For public data to have real value it needs to be easily discoverable, accessible and usable (e.g. de La Beaujardière et al. 2010), which means:

- Detailed information about data quality and provenance are recorded;
- Data and metadata are obtainable in standard formats, so that they can be imported into commonly used analysis tools;
- Tools are available to help identify data of interest among all the public data available.

Infrastructure for data management and access

The challenges of making diverse data collections accessible and useable to diverse user groups are common to

all Earth Sciences. In response, systems have been built around broad themes or disciplines, offering integration, standardisation and tools.

The European EarthServer project (Baumann et al. 2015) is focused on “Big Data Analytics”, providing interoperable access, visualisation and processing services for data collections that are large in volume, complex, heterogeneous and often growing rapidly. The infrastructure is based on data stored in the *rasdaman*¹ array database management system, with retrieval and analysis via Web Coverage Service and Web Coverage Processing Service (WCS and WCPS, standards of the Open Geospatial Consortium² (OGC)).

The Data Observation Network for Earth³ (DataONE) is a US-based project to improve international cross-disciplinary data access in the biological, environmental and Earth sciences (Michener et al. 2011; Cohn 2012). It provides centralised access to distributed data, tools for data discovery and analysis, and persistent identifiers to enable citation of digital objects.

The Ocean Biogeographic Information System’s Spatial Ecological Analysis of Marine-megavertebrate Animal Populations (OBIS-SEAMAP, Halpin et al. 2009) provides data management services, data integration and data query and visualisation tools for researchers studying marine species. The system is based on open standards and open-source software.

In the marine domain progress in establishing these information infrastructures to underpin ocean observing systems is occurring world-wide at varying levels. In Europe the leading example is the collection of data portals under EMODnet,⁴ where each discipline has its own portal for collating Europe-wide data; Europe also has regional implementations e.g. COSYNA⁵ in the German Bight and SOCIB⁶ for the Balearics in Spain. In North America Canada has the OCEAN NETWORKS CANADA⁷ and the US has IOOS.⁸ OBIS, EMODnet, SOCIB, US-IOOS also make use of OGC-standard web services such as Web Map Service and Web Feature Service (WMS & WFS).

These examples illustrate the significant advances that are being made in developing information infrastructures. Although no single system serving all Earth Sciences has been globally adopted to date, there is some convergence towards interoperability among these projects, in particular through the use of web services and adoption of OGC standards.

Australia’s Integrated Marine Observing System

Australia has the third-largest marine jurisdiction of any nation on Earth, 13.86 million km²; Australia’s Search and Rescue region extends over 52.8 million km², one tenth of the Earth’s surface. In contrast, 85 % of the population live within 50 km of the coast. Now in its 10th year, the Australian Integrated Marine Observing System (IMOS) was established to address the requirements of the marine and climate science community, and contribute to international ocean observing programs (Proctor et al. 2010; Hill et al. 2010; Seim et al. 2010; Lynch et al. 2014). IMOS, with a focus on sustained observing of the ocean, is a component of the National Collaborative Research Infrastructure Strategy (NCRIS)⁹ funded by the Department of Education of the Australian Government. IMOS receives core funding from NCRIS, with co-investment and in-kind support from a range of universities, research institutes, state governments and other organisations. IMOS is considered sufficiently mature to be recognised as a member of the GOOS (Global Ocean Observing System) Regional Alliance,¹⁰ one of 13 Regional Alliances covering the greater part of the world’s oceans and marginal seas.

IMOS is integrated in terms of its geographic domain, from coast to open ocean, and scientific domain, combining physical, chemical and biological variables. The observing program is guided by regional science plans developed by the research community, focusing on the major themes of multi-decadal ocean change, climate variability and weather extremes, major boundary currents and inter-basin flows, continental shelf processes, and ecosystem responses (<http://imos.org.au/nodes.html>).

The observations identified in these regional science plans are carried out by national facilities, operated by partner institutions around the country, each facility deploying a particular type of observing platform. Data holdings, as of mid-2015 include:

- Argo floats (~800 floats routinely in the Australian region, ~300,000 profiles to date),
- Ships of opportunity (air-sea flux/SST: 3 vessels, CO₂: 4, bio-acoustics: 10, continuous plankton recorder: 11, XBT: 30 lines),
- Deep-water moorings (11 platforms, 23 deployments),
- Ocean gliders (26 platforms, 170 deployments),
- Autonomous underwater vehicles (10 sites, 35 deployment campaigns, >3 million images),
- Shelf/coastal moorings (>50 sites, ~1300 deployments)

¹ <http://www.rasdaman.org>

² <http://www.opengeospatial.org>

³ <https://www.dataone.org>

⁴ <http://www.emodnet.eu>

⁵ http://www.hzg.de/institutes_platforms/cosyna/index.php.en

⁶ <http://www.socib.eu>

⁷ <http://www.oceannetworks.ca>

⁸ <http://www.ioos.noaa.gov>

⁹ <https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncris>

¹⁰ http://www.ioc-goos.org/index.php?option=com_content&view=article&id=159&Itemid=89

- Coastal ocean radars (6 sites, up to 8 years of hourly gridded velocity products),
- Animal tracking and monitoring (~230,000 CTD profiles from satellite tags, ~70 million acoustic detections),
- Wireless sensor network (7 sites, 290 sensors),
- Satellite remote sensing (>20 years of daily SST & altimetry products, ~13 years of daily ocean colour products).

During IMOS, and pre-IMOS, a number of Observation Simulation Experiments (OSEs) have been undertaken to assess the efficacy of IMOS observing platforms to capture a representative variation of ocean behaviour or to explore possibilities for expanding, or reformulating, the observing system. For example Oke et al. (2009) considered various options for elements of an observing system along the coast of New South Wales, Australia, in terms of their benefits to an ocean forecast and reanalysis system. They assessed the likely benefits of assimilating in situ temperature (T) and salinity (S) observations from repeat glider transects and surface velocity observations from high-frequency (HF) radar arrays into an eddy-resolving ocean model. The study demonstrated that if HF radar observations are assimilated along with the standard components of the global ocean observing system, the analysis errors are likely to reduce by as much as 80 % for velocity and 60 % for T, S and sea-level in the vicinity of the observations. Owing to the relatively short along-shore decorrelation length-scales for T and S near the shelf, the glider observations are likely to provide the forecast system with a more modest gain.

The assessment of the physical components of the National Reference Stations (NRS, a component of the coastal moorings facility) carried out by Oke and Sakov (2012) identified that in combination, the nine NRSs effectively monitor the inter-annual variability of the continental shelf circulation in about 80 % of the region around Australia. Other studies that have used NRS data in part to investigate oceanographic phenomena, such as the 2011 marine heat wave in Western Australia (Feng et al. 2013), show at least for some parameters, the NRS can monitor and detect large scale patterns, events and anomalies rather than just local events, exceeding design expectations.

The value of IMOS observations is demonstrated by its impact on a wide range of research activities. As of February 2016, the IMOS program and its observations have been cited in over 580 journal articles, over 1500 conference and workshop presentations, and over 400 other publications (including reports, white papers and theses). IMOS data have been used in over 190 postgraduate research projects, over 300 other research projects and incorporated in a range of data products (detailed listings can be found at <http://imos.org.au/imospublications.html>).

IMOS provides the single most important ocean observing information infrastructure in Australia; it is the primary source

of ocean observations for the National Environmental Information Infrastructure¹¹, a core activity under the Australian Government's National Plan for Environmental Information initiative.¹² Where ever possible efforts have been made to align these infrastructures by utilising the tools and standards described below.

The IMOS information infrastructure

The facilities deliver the observations and associated metadata to the IMOS data centre which, through controlled workflows, conducts assessment and archival, and provides infrastructure (Fig. 1) for the delivery of the data to the research community and the public. Inherent in this goal is the aim to provide consistency in data quality, formats, metadata, and interoperability with other programs and data sources. The approach to this goal, as described in this paper, is through adoption of international standards (e.g. OGC), development of common data processing and compliance checking tools, contributions to open source software (e.g. GeoNetwork Opensource¹³), and participation in international projects (e.g. Ocean Data Interoperability Platform, ODIP¹⁴).

The design of the infrastructure was guided by the requirements of our users, practical constraints, the goal of interoperability, and the diverse nature of the observations. In particular, the infrastructure needs to a) be robust and scalable; b) provide a central point of discovery and access to IMOS data; c) handle a wide range of physical, chemical and biological observations from a variety of platforms; d) handle a range of data “shapes” (e.g. timeseries, gridded data, ship tracks and profiles); e) be interoperable with other programs and data sources; and f) serve the needs of data users from diverse backgrounds and with various levels of computing expertise.

In the sections below we describe the main components of the infrastructure and workflows we have set up to meet the above requirements. Many of the systems and processes we use have changed significantly since previously published descriptions, and continue to evolve. Finally we outline current and planned developments, including the use of this infrastructure in a broader context.

Formats and conventions

We have adopted the widely used Network Common Data Form (NetCDF, Unidata 2015) as the primary format for transfer and archival of data within IMOS, and for delivering data to users. Using NetCDF has many advantages, it is self-

¹¹ <http://www.neii.gov.au/>

¹² <http://www.bom.gov.au/environment/about.shtml>

¹³ <http://geonetwork-opensource.org/>

¹⁴ <http://www.odip.eu>

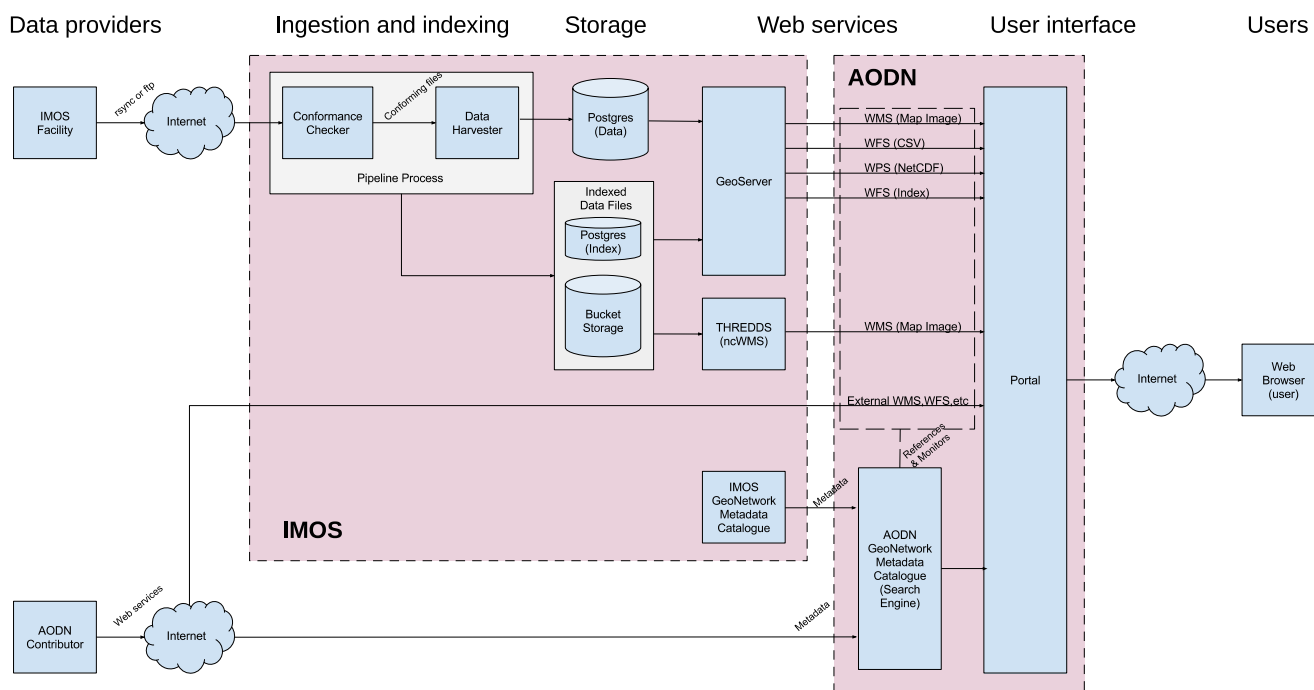


Fig. 1 Main components of the IMOS information infrastructure, illustrating how it transitions to the Australian Ocean Data Network (AODN) by including data from sources outside IMOS. All components

are open-source and freely available to contributors to assist in providing standard web services and metadata

describing (i.e. stores data and metadata together), flexible, able to store large volumes of multi-dimensional data, and can be read by common data-analysis tools. For interoperability with available analysis tools, we require files to be structured and documented according to the Climate and Forecasting conventions¹⁵ (CF, version 1.6). These conventions are commonly used in oceanography and are designed to promote the processing and sharing of NetCDF files. As described on the CF website, the metadata defined by the conventions “enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, re-gridding, and display capabilities.” The NetCDF format and the CF conventions are now OGC standards. We also specify additional, IMOS-specific conventions (IMOS 2015) that make data management and discovery easier.

We have adopted the open-source IOOS Compliance Checker¹⁶ package developed by the US Integrated Ocean Observing System as a tool to validate NetCDF files against the CF conventions. This was chosen because it is written in Python (therefore easily modified), it already had CF checks built in, it is easily extensible with plug-ins, and it was a welcome opportunity to collaborate with the US-IOOS developers. We are contributing to further development of this package, and adding IMOS-specific check suites based on the same framework.

Whilst, due to its self-describing nature, NetCDF is ideal for archiving and transferral of data, the IMOS portal aims to offer high granularity. This is achieved by harvesting the data and metadata into a database and enabling the user to make web service calls (Map service, Feature service) based on faceted subsetting at differing levels of granularity depending on the data collections selected (see below). This gives the user access to data at the level of their choosing.

Data processing and quality control

IMOS data are collected and processed by numerous operators around the country. In some cases, data from distinct platforms of the same type (e.g. coastal moorings) are provided by different institutions. To ensure consistency of data quality and format, we have developed the IMOS Matlab Toolbox,¹⁷ a standard tool to

1. Read data in various instrument-specific formats;
2. Read associated metadata from a database;
3. Perform conversions, corrections and compute derived variables;
4. Apply quality-assurance/quality-control (QA/QC) tests; and
5. Write NetCDF files according to the required conventions.

¹⁵ <http://cfconventions.org/>

¹⁶ <https://github.com/ioos/compliance-checker>

¹⁷ <https://github.com/aodn/imos-toolbox>

The design of the Toolbox was guided by a number of factors: Matlab is commonly used by the IMOS Facility operators; some of them were already using a deployment database; IMOS deploys many different instruments, hence the need for a range of parsers. The processing and QA/QC procedures implemented in the Toolbox¹⁸ are continually developed in collaboration with the IMOS facility operators and other experts (e.g. Morello et al. 2014). Every data point carries a QC flag. The Toolbox data flow schema is shown in Fig. 2.

At the time of writing the Toolbox supports processing of multiple sensor data from Seabird, WET labs, Teledyne, Nortek, and others (see <https://github.com/aodn/imos-toolbox/wiki/SupportedInstruments>). The Toolbox is available as a standalone binary running under the Matlab Component Runtime (MCR); as well as a full installation version requiring Matlab version R2012b or later (both fully documented at <https://github.com/aodn/imos-toolbox/wiki>).

Data ingestion

IMOS data are typically provided to the data centre in the form of NetCDF files, uploaded via file transfer protocol (FTP). We use FTP because it is simple and easy to use, and more importantly, it saves us the cost of developing and maintaining a specific application to manage data uploads. We also receive data in other formats, including comma-separated values (CSV), Microsoft Access® databases, or as web services.

Data ingestion is carried out by event-driven “pipeline” processes, triggered by any new file that appears in a designated incoming directory (Fig. 3). The file is validated against the required conventions and, provided it passes all the checks, moved to publicly available storage. If the file is not compliant, it is moved to an error directory and details of the failed checks are sent to the data provider. Additional processing steps, such as converting data from other formats into NetCDF or adding metadata, can easily be included in the pipeline. If data are processed using the IMOS Matlab Toolbox compliance is guaranteed.

Detailed metadata (e.g. spatial and temporal extent, platform and instrument details, units of measure and data quality information) are stored in each self-describing NetCDF file. These metadata are harvested from all published NetCDF files into a PostgreSQL/PostGIS¹⁹ database using Talend Open Studio for Data Integration²⁰ (Extract, Transform, Load software). Storing this information in a database enables the use of complex queries for collection filtering and reporting on data

availability, for example to select only files that contain data within a specified geographic bounding box and a time interval. Having the metadata in a database also enables us to create web services (described in the next section).

For non-gridded files (generally one-dimensional or small two-dimensional data sets) the data are also harvested into the database. This allows fine-grained data subsetting, i.e. selecting only data points that meet user-specified constraints (based on location, time or the value of any measured parameter). Further, it allows direct data access via WFS, which in turn provides a range of download formats (in particular the much-requested CSV). Having the metadata and data in the database gives us complete flexibility over the granularity of the available information. For large gridded data sets (e.g. satellite products, HF radar products) we do not harvest data into the database for space and performance reasons. However, we have other tools to handle subsetting & aggregation of these data (see *Download* section below).

This ingestion process inherently groups data into distinct collections. In this context, we define a “collection” as the largest set of data that can be viewed on a single map layer or downloaded in one request. As we are transferring data from NetCDF files to a database, the database schema needs to replicate the structure of the files. For performance reasons, the data in the database are stored in flat tables, with specific columns for specific measured parameters (e.g. temperature or phytoplankton abundance). This flat structure is also needed so that we can create a WFS and output a plain-text (CSV) version of the data, as required by users. Thus each collection consists of files with the same structure and set of parameters. For example we have a collection of temperature and salinity time-series from moorings, and another collection for CTD profiles.

Data discovery and access

Metadata

Data collections are described by metadata records structured according to the Marine Community Profile²¹ (MCP v2.0), a subset and extension of the ISO 19115 standard (ISO 2003), including links to specific map and download services. The MCP was developed several years ago because the ISO 19115 standard was considered too generic for oceanography whilst omitting elements (e.g. measured parameters) important to the discipline. To facilitate data discovery by faceted search, measured parameters, units of measure, platforms, instruments and organisations are specified using terms from the Australian Ocean Data Network (AODN) controlled

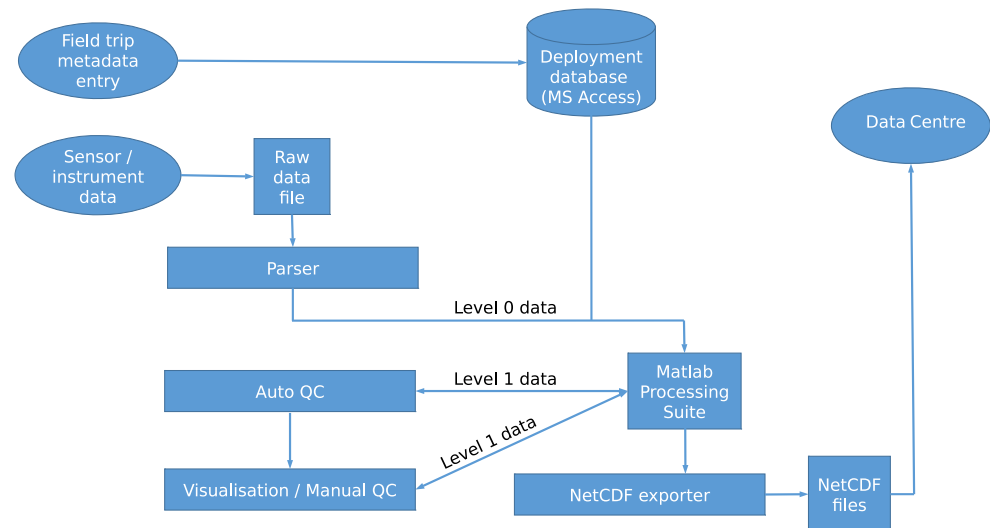
¹⁸ <https://github.com/aodn/imos-toolbox/wiki/QCProcedures>

¹⁹ <http://www.postgresql.org/>; <http://postgis.net/>

²⁰ <http://www.talend.com/products/talend-open-studio>

²¹ <http://bluenet3.antcrc.utas.edu.au/mcp-doc-20-draft/>

Fig. 2 Data flow schema for IMOS Toolbox for processing sensor data. Level 0 data is unprocessed (raw) data; Level 1 data is Quality-Controlled data



vocabularies (accessible through the Australian National Data Service at <https://vocabs.ands.org.au/#/?q=AODN>). Faceted (or navigational) search uses a hierarchy structure (taxonomy) to enable users to browse information by choosing from a pre-determined set of categories. This allows a user to refine their search options by navigating/drilling down a category. It is an advanced search, made easier for the user by the visible folder structure. Examples of this can be seen on eCommerce sites like Amazon and eBay.

The metadata records are managed and accessed using GeoNetwork Opensource software. We have contributed code to this project to enable the use of hierarchical vocabulary terms in a faceted search; and to temporarily hide from public view any metadata record with linked web services that are currently unavailable. The instance of GeoNetwork which supports the IMOS Ocean Portal is available at <https://catalogue-portal.aodn.org.au>.

Mapping

Maps showing the spatial distribution of data collections are created based on the metadata harvested into the database; all data collections contain geospatial information and this is included in the metadata in the NetCDF files (and contributes to their self-description). The maps are published using GeoServer²² as Web Map Services (WMS; de La Beaujardière 2002), a standard of the Open Geospatial Consortium (OGC). Geoserver is an active open-source project that does most of what we need (WMS, WFS, filtering). For gridded data sets, maps are generated directly from the NetCDF files using a WMS application, ncWMS²³ (Blower et al. 2013), specifically designed for use with NetCDF files.

The maps are dynamically generated to reflect the subset specified by the user.

Download

Data in the database are served as OGC-standard Web Feature Services (WFS; Vretanos 2005) delivered by GeoServer, allowing subsets to be downloaded in a range of formats including comma-separated values (CSV), Geography Markup Language (GML), Keyhole Markup Language (KML), and JavaScript Object Notation (JSON). Due to its simplicity and compatibility with spreadsheet software, CSV is the most commonly requested format by our users. We have created a GeoServer plug-in that adds a basic metadata header to the CSV output, providing information about column names and quality control flags. All WMS/WFS services are monitored by Nagios²⁴ for their ‘live’ status, and alerts issued should a problem be detected.

We also provide WFS services with coarser granularity, where each feature represents one complete NetCDF file (as received from the IMOS facility operators). This allows the list of published files to be searched based on spatial/temporal extent and other attributes of each file, returning a download URL for each selected file.

For gridded data, we have developed a simple subset and aggregation service (GoGoDuck²⁵) based on the NetCDF Operators (NCO) toolkit,²⁶ to generate NetCDF files for download. Work is in progress to provide similar functionality for non-gridded data (e.g. time-series, trajectories, profiles) using the OGC Web Processing Service (WPS, Schut and Whiteside 2007) interface provided by GeoServer. Data matching the user’s subset criteria is extracted from the

²² <http://geoserver.org/>

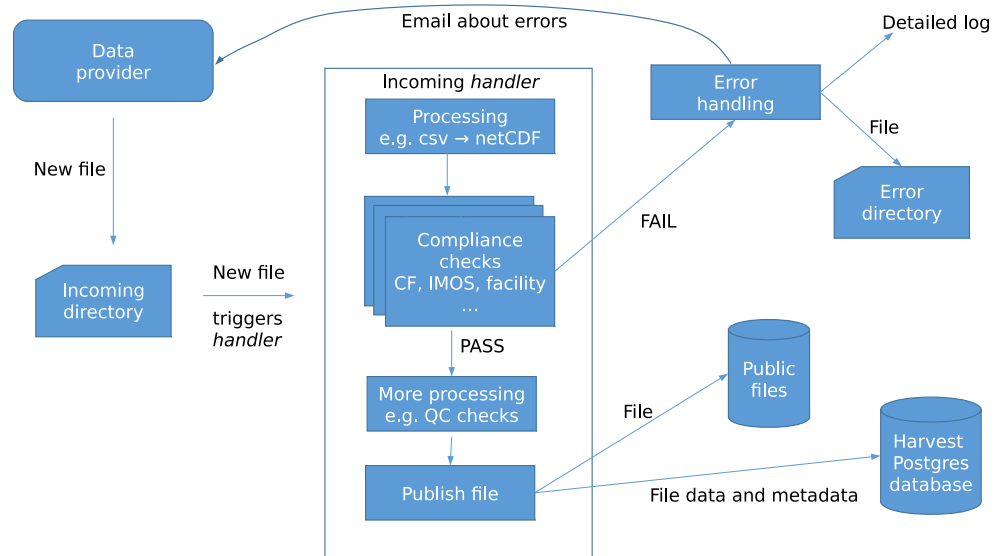
²³ <http://ncwms.sf.net>

²⁴ <https://www.nagios.org/>

²⁵ <https://github.com/aodn/go-go-duck>

²⁶ <http://nco.sourceforge.net/>

Fig. 3 Structure of the event-driven data ingestion pipelines being set up to automatically process files from IMOS data providers as soon as they are uploaded to an incoming directory



database and converted into NetCDF files based on a collection-specific file template. We are providing these options because some users want to download subsets and/or aggregations of files, or individual variables within files, in NetCDF format. This can reduce download size and provide more practical access than obtaining all the original NetCDF files.

In addition, all IMOS data files can be downloaded via a simple web-based data server (<http://data.aodn.org.au/>). NetCDF files can also be browsed and downloaded via a THREDDS Data Server²⁷ (<http://thredds.aodn.org.au>) or accessed remotely via the OPeNDAP²⁸ protocol. THREDDS and OPeNDAP are often the preferred option for advanced users who routinely access large numbers of files.

User Interface

It is possible to discover and access IMOS data directly through the above services. However, to make this process easier and quicker, we have developed the Open Geospatial Portal,²⁹ an open-source software project, to provide a unified web interface. The need for this portal was to meet the constraints and requirements from our users, notably to discover and access data from many data providers (the IMOS facilities), with many parameters & products of physical, chemical, biological data, and recognising that many users only want Excel/CSV files, whilst more advanced users want lots of data at once and are able to use NetCDF, THREDDS, and web services. The interface (<https://imos.aodn.org.au>) enables a user to explore the IMOS collections in three simple steps: (1) search, (2) subset and preview, and (3) download.

²⁷ <http://www.unidata.ucar.edu/software/thredds/current/tds/TDS.html>

²⁸ <http://www.opendap.org/>

²⁹ <https://github.com/aodn/aodn-portal>

Search

Step 1 relies on the content of metadata records. It allows users to identify collections of interest by specifying a geographic boundary, date range, measured parameters, platform types or organisations responsible for the data. The parameters, platform types and organisations are selected from the AODN controlled vocabularies. A free-text search of all metadata fields is also possible. Once choices have been made the metadata records of data collections meeting the criteria are presented and any, or all, of these collections can be selected for subsetting in Step 2.

Subset and preview

Step 2 uses the capabilities of GeoServer and ncWMS to display a map (i.e. WMS) showing the geographic coverage of the selected data collections. An example is shown in Fig. 4. The WMS layers can then be filtered according to spatial, temporal and other constraints specific to each collection (e.g. vessel names or availability of a given parameter), and the map is updated to reflect the spatial extent of the data subset that will be downloaded. For gridded data products the map gives a preview of the actual data for a selected parameter and time slice, and allows backward and forward time stepping. The interface also gives access to the abstract and links to online resources for each selected collection.

Download

In Step 3, based on their requests, users are offered a choice of formats to download the selected subsets. A direct URL link to each selected collection is also given for future reference or sharing (the link opens the portal

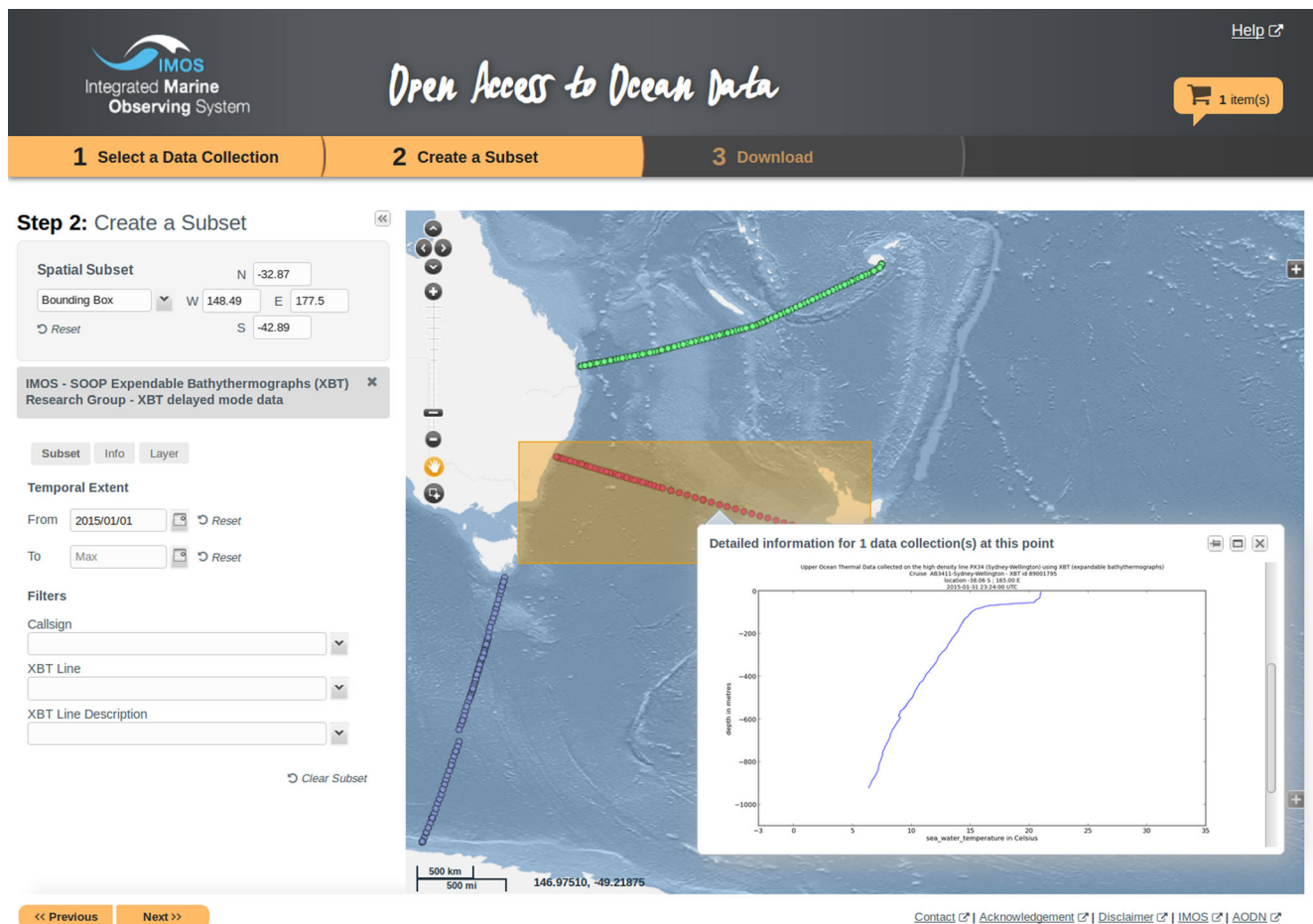


Fig. 4 Screenshot of the Open Geospatial Portal showing Step 2, with a collection of XBT data from ships of opportunity selected. The spatial, temporal and metadata filters in the left panel are used to select a subset of

the data, which is previewed on the map and available for download in Step 3. Clicking on a feature in the map provides additional information about that feature

on step 2, with the given collection already selected). Currently the available download formats are

- The original NetCDF files that contribute to the selected subset (no subsetting within each file);
- A single, subset and aggregated NetCDF file (for gridded data only);
- A list of URL links to the original NetCDF files;
- A text file of comma-separated values (CSV), including basic metadata;
- A Python code snippet to access the data directly via a WFS query.

Specialised data access tools

Whilst all IMOS data collections can be visualised on a map, and details of installations, deployments, etc. readily downloaded via the Ocean Portal, some data products are better served through specialised tools providing

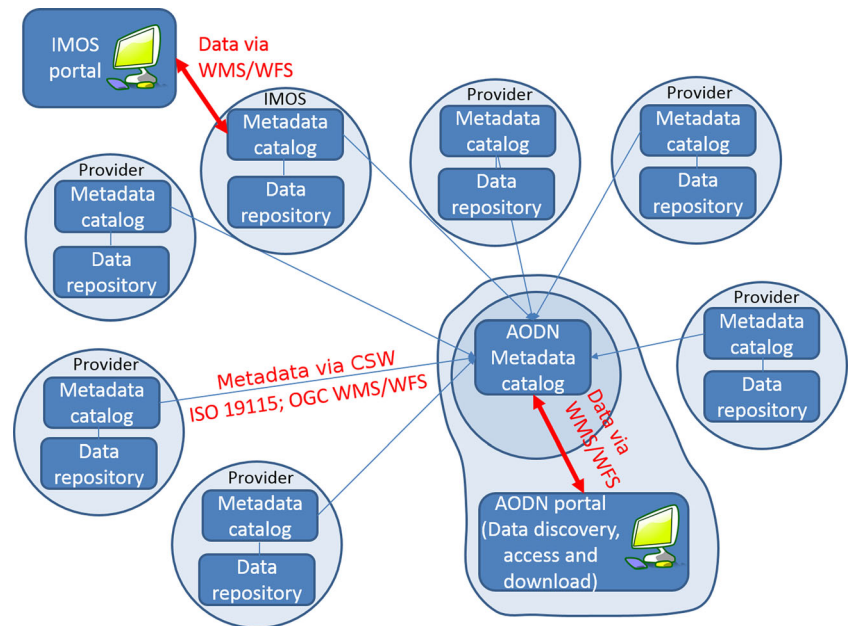
additional functionality. These tools are accessible through the Portal. For example:

- The acoustic detection data from the Australian Animal Tracking and Monitoring System (<http://imos.org.au/animaltracking.html>) is available through a web application, <https://aatams.emii.org.au>, described in GitHub,³⁰ which enables contributors to upload data and users to download data. Due to the sensitive nature of some of the tagged species, different levels of ‘security’ are imposed on downloads (e.g. fuzzy locations) according to user status.
- Images produced along tracks surveyed by the Automated Underwater Vehicles facility (<http://imos.org.au/auv.html>) are better accessed through a dedicated viewer,³¹ <https://auv.aodn.org.au/auv/>, which provides faster pan and zoom, preview of

³⁰ <https://github.com/aodn/aatams>

³¹ <https://github.com/aodn/auv>

Fig. 5 Schematic for the distributed information infrastructure of the Australian Ocean Data Network (AODN)



multiple images with associated metadata, and image download features.

- Sound spectra from recordings produced by the Passive Acoustic Observatories of the National Mooring Network (<http://imos.org.au/nationalmooringnetwork.html>) are more readily manipulated by a special viewer³² (<https://acoustic.aodn.org.au>)
- Three-dimensional visualisation of glider data is better viewed through Gliderscope,³³ a desktop package developed by the Australian National Facility for Ocean Gliders (<http://imos.org.au/oceangliders.html>).

These, and other tools, are also available directly from the ‘IMOS Tools’ page on the IMOS website (<http://imos.org.au/imosdatatools.html>).

Summary and future work

We have described the infrastructure we use to manage and distribute data collected by Australia’s Integrated Marine Observing System. We have developed an information infrastructure based on open-source tools, which employs controlled vocabularies, and utilises recognised (e.g. OGC) data and web services, and which has the power to deliver comprehensive metadata and data with granularity ranging from a data collection(s) down to individual data elements through an intuitive-to-use, stepwise, geospatial portal, with options to access data in other conventional ways. We believe there are

few infrastructures in the marine domain which deliver this data granularity across multiple disciplines.

The same infrastructure can be adapted to work in a broader context, serving a diverse range of observations from multiple data providers. Indeed work is in progress to make data sets collected by other organisations (not funded by IMOS) available via the IMOS Ocean Portal, with the intention that this portal will become the Australian Ocean Data Network³⁴ (AODN). This is indicated in Fig. 1. Other organisations serving their own data collections have also set up infrastructure using the same open-source components (e.g. the Institute of Marine and Antarctic Studies Data Portal³⁵) enabling them to retain institutional identity whilst delivering interoperable data collections. Thus a distributed network of data infrastructures can be readily integrated and accessed through a single point of entry, i.e. AODN, as illustrated in Fig. 5.

Indeed, the recently published Australian National Marine Science Plan (2015) recognises that a fully embracing, federated and standards-based AODN will give Australian marine science a competitive edge.

We will continue to develop the infrastructure, focusing on improving reliability and ease of access. Current developments include:

- Expanding the controlled vocabularies used in the faceted search on the portal;
- Improving data consistency and compliance by introducing an automated compliance checker;

³² <https://github.com/aodn/acoustic-data-viewer>

³³ <http://imos.org.au/gliderscope.html>

³⁴ <http://imos.org.au/aodn.html>

³⁵ <http://data.imas.utas.edu.au/>

- Adding functionality to display multiple map layers from the same collection;
- Adding more facets to the portal search page;
- Documenting in detail the work required to set up our infrastructure
- Improving quality control processes implemented in the IMOS Matlab Toolbox;

It is also our policy to keep abreast of latest developments in the open source communities, e.g. GeoNetwork, GeoServer. We welcome contributions to the Open Geospatial Portal project, and any other open-source software it relies on.

Acknowledgments IMOS is a national collaborative research infrastructure, supported by Australian government. It is led by University of Tasmania in partnership with the Australian marine & climate science community.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baumann P, Mazzetti P, Ungar J, Barbera R, Barboni D, Beccati A, *et al.* (2015). Big Data Analytics for Earth Sciences: The EarthServer Approach. *International Journal of Digital Earth*, (ahead-of-print), 1–27. doi:10.1080/17538947.2014.1003106
- Blower JD, Gemmell AL, Griffiths GH, Haines K, Santokhee A, Yang X (2013) A web map service implementation for the visualization of multidimensional gridded environmental data. *Environ Model Softw* 47:218–224
- Cohn JP (2012) DataONE opens doors to scientists across disciplines. *Bioscience* 62(11):1004–1004. doi:10.1525/bio.2012.62.11.16
- de La Beaujardière J (2002) Web Map Service Implementation Specification, Version 1.1.1, OGC 01-068r3. http://portal.opengeospatial.org/files/?artifact_id=1081&version=1&format=pdf. Accessed 30 September 2015
- de La Beaujardière J, Beegle-Krause C, Bermudez L, Hankin S, Hazard L, Howlet E, Le S, Proctor R, Signell RP, Snowden D, Thomas J (2010) Ocean and Coastal Data Management. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 1), Venice, Italy, 21–25 September 2009
- Feng M, McPhaden MJ, Xie S-P, Hafner J (2013) La Niña forces unprecedented Leeuwin current warming in 2011. *Sci Rep* 3:1277. doi:10.1038/srep01277
- Halpin PN, Read AJ, Fujioka E, Best BD, Donnelly B, Hazen LJ, Kot C, Urian K, LaBrecque E, Dimatteo A, Cleary J, Good C, Crowder LB, Hyrenbach KD (2009) OBIS-SEAMAP: the world data center for marine mammal, sea bird, and sea turtle distributions. *Oceanography* 22(2):104–115. doi:10.5670/oceanog.2009.42
- Hill K, Moltmann T, Proctor R, Allen S (2010) The Australian integrated marine observing system: delivering data streams to address national and international research priorities. *Mar Technol Soc J* 44(6):65–72
- IMOS (2015) IMOS NetCDF Conventions Available at http://help.aodn.org.au/help/sites/help.aodn.org.au/files/IMOS_NetCDF_Conventions.pdf. Accessed 30 September 2015
- ISO (2003) 19115:2003 Standard (Geographic information – Metadata). Available at <http://www.iso.org/iso/home.html>
- Lynch T, Morello EB, Evans K, Richardson A, Steinberg C, Roughan M, Thompson P, Middleton J, Feng M, Brando V, Tilbrook B, Ridgway K, Allen S, Doherty P, Hill K, Moltmann T (2014) IMOS National Reference Stations: a continental scaled physical, chemical and biological coastal observing system. *PLoS One* 9(12):e113652
- Michener W, Viegais D, Vision T, Kunze J, Cruse P, Janée G (2011) DataONE: data observation network for earth-preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine* 17(1):3. doi:10.1045/january2011-michener
- Morello EB, Galibert G, Smith D, Ridgway KR, Howell B, Slawinski D, Timms GP, Evans K, Lynch TP (2014) Quality control (QC) procedures for Australia's National Reference Stations: a comparing semi-autonomous systems to an expert oceanographer. *Methods in Oceanography* 9:17–33
- National Marine Science Plan (2015–2025): Driving the development of Australia's blue economy. www.marinesciencenet.au August 2015. 47 pp.
- Oke PR, Sakov P (2012) Assessing the footprint of a regional ocean observing system. *J Mar Syst* 105–108:30–51. doi:10.1016/j.jmarsys.2012.05.009
- Oke PR, Schulz E (2009) A comparison of shelf observation platforms for assimilation in an eddy-resolving ocean model. *Dynamics of Atmospheres and Oceans* 48:121–142
- Proctor R, Roberts K, Ward BJ (2010) A data delivery system for IMOS, the Australian integrated marine observing system. *Adv Geosci* 28: 11–16
- Schut P & Whiteside A (2007) OpenGIS® Web Processing Service, OpenGIS® Standard, Version 1.0.0, OGC 05-007r7. http://portal.opengeospatial.org/files/?artifact_id=24151. Accessed 30 September 2015
- Seim HE, Dahlin H, Meyers G, Shuford R, Proctor R (2010) Development of Delivery of Services From Ocean Observing Systems - An Opportunity to Promote Common Approaches for a Global Ocean Observing System. *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 1), Venice, Italy, 21–25 September 2009
- Unidata (2015) Network common data form (NetCDF). Boulder, CO: UCAR/Unidata. doi: 10.5065/D6H70CW6. Accessed 30 September 2015
- Vretanos P A (2005) Web feature service implementation specification. Version 1.1.0, OGC 04–094. http://portal.opengeospatial.org/files/?artifact_id=8339. Accessed 30 September 2015